How to think scientifically about scientists' proposals for fixing science[*]

Andrew Gelman[†]

10 Feb 2017

Science is in crisis. Any doubt about this status has surely been been dispelled by the loud assurances to the contrary by various authority figures who are deeply invested in the current system and have written things such as, "Psychology is not in crisis, contrary to popular rumor . . . Crisis or no crisis, the field develops consensus about the most valuable insights . . . National panels will convene and caution scientists, reviewers, and editors to uphold standards." (Fiske, Schacter, and Taylor, 2016). When leaders go to that much trouble to insist there is no problem, it's only natural for outsiders to worry.[1]

The present article is being written for a sociology journal, which is appropriate for two reasons. First, sociology includes the study of institutions and communities; modern science is both an institution and a community, and as such it would be of interest to me as a citizen and a political scientist, even beyond my direct involvement as a practicing researcher. Second, sociology has a tradition of questioning; it is a field from whose luminaries I hope never to hear platitudes such as "Crisis or no crisis, the field develops consensus about the most valuable insights." Sociology, like statistics and political science, is inherently accepting of uncertainty and variation. Following Karl Popper, Thomas Kuhn, Imre Lakatos, and Deborah Mayo, we cheerfully build our theories as tall and broad as we can, in the full awareness that reality will knock them down. We know that one of the key purposes of data analysis is to "kill our darlings," and we also know that the more specific we make our models, the more we learn from their rejection. Structured modeling and thick description go together.

Just as we learn in a local way from our modeling failures, we can learn more globally from crises in entire subfields of science. When I say that the replication crisis is also an opportunity, this is more than a fortune-cookie cliche; it is also a recognition that when a group of people make a series of bad decisions, this motivates a search for what went wrong in their decision-making process.

A full discussion of the crisis in science would include three parts:

1. Evidence that science is indeed in crisis: at the very least, a series of examples of prominent products of mainstream science that were seriously flawed but still strongly promoted by the scientific community, and some evidence or at least speculation that such problems are prevalent enough to be worth our concern.

2. A discussion of what has gone wrong in the ideas and methods of scientific inquiry and in the process by which scientific claims are promoted and disseminated within the community and the larger society. This discussion could include specific concerns about statistical methods

---

[*]For a special issue of *Socius* on Fixing Science, ed. David Grusky, Lisa Keister, and James Moody.

[†]Department of Statistics and Department of Political Science, Columbia University.

[1]At this point a savvy critic might point to global-warming denialism and HIV/AIDS denialism as examples where the scientific consensus is to be trusted and where the dissidents are the crazies and the hacks. Without commenting on the specifics of these fields, I will just point out that the research leaders in those areas are *not* declaring a lack of crisis—far from it!—nor are they shilling for their "patterns of discovery." Rather, the leaders in these fields have been raising the alarm for decades and have been actively pointing out inconsistencies in their theories and gaps in their understanding. Thus, I do not think that my recommendation to watch out when the experts tell you to calm down, implies blanket support for dissidents in all areas of science. One's attitude toward dissidents should depend a bit on the openness to inquiry of the establishments from which they are dissenting.

such as null hypothesis significance testing, and also institutional issues such as the increasing pressure on research to publish large numbers of articles.

3. Proposed solutions, which again range from research methods (for example, the suggestion to perform within-person, rather than between-person, comparisons wherever possible) to rules such as preregistration of hypotheses, to changes in the system of scientific publication and credit.

I and others have written enough on topics 1 and 2, and since this article has been solicited for a collection on Fixing Science, I'll restrict my attention to topic 3: what to do about the problem?

If you've gone to the trouble to pick up (or click on) this volume in the first place, you've probably already seen, somewhere or another, most of the ideas I could possibly propose on how science should be fixed. My focus here will not be on the suggestions themselves but rather on what are our reasons for thinking these proposed innovations might be good ideas. The unfortunate paradox is that the very aspects of "junk science" that we so properly criticize—the reliance on indirect, highly variable measurements from nonrepresentative samples, open-ended data analysis, followed up by grandiose conclusions and emphatic policy recommendations drawn from questionable data—all seem to occur when we suggest our own improvements to the system. All our carefully-held principles seem to evaporate when our emotions get engaged. This is similar to a pattern noted by Gelman and Loken (2012) that academic statisticians only rarely seem to use statistical principles in designing and evaluating their teaching.

I will now discuss various suggested solutions to the replication crisis, and the difficulty of using scientific evidence to guess at their effects.

The first set of reforms are *educational*: beef up college statistics requirements; improve textbooks and online teaching materials on research methods and statistics to better educate the students of the future; set up mid-career initiatives to bring working scientists up to speed; and write blogs and news articles to spread awareness of these opportunities. These may well be excellent ideas, but what evidence is there that they will "fix science" in any way. Given the widespread misunderstandings of statistical and research methods, even among statisticians, what makes us so sure that more classroom or training hours will make a difference?

The second set of reforms involve *statistical methods*: reduce reliance of p-values and hypothesis testing, instead encouraging alternatives such as confidence intervals, Bayes factors, multilevel models, exploratory data analysis, and nonparametric methods—to name just a few of the proposed alternatives. I am a loud proponent of some of these ideas, but, again, I come bearing no *statistical* evidence that they will improve scientific practice. My colleagues and I have given many examples of modern statistical methods solving problems and resolving confusions that arose from null hypothesis significance testing, and our many good stories in this vein represent . . . what's the plural of "anecdote," again?

A related set of ideas involve revised *research practices*, including open data and code, preregistration, and, more generally, a clearer integration of workflow into scientific practice. I favor all these ideas, to different extents, and have been trying to do more of them myself. But, again, I don't see the data demonstrating their effectiveness. If the community of science critics (to which I consider myself a member) were to hold the "open data" movement to the same standards that we demand for research such as "power pose," we would have no choice but to label all these ideas as speculative.

The third set of proposed reforms are *institutional* and involve altering the existing incentives that favor shoddy science and that raise the relative costs to doing good, careful work. Suggestions here include judging work by quality rather than quantity, disregarding "impact factors" for scientific journals, changing promotion and tenure criteria, requiring scientists to take responsibilities for press releases about their work, and, my personal favorite, open publishing followed by post-publication review. All these proposals sound good to me. But, again, no evidence.

The foregoing review is intended to be thought provoking, but not nihilistic. One of the most important statistical lessons from the recent replication crisis is that certainty or even near-certainty is harder to come by then most of us had imagined. We need to make some decisions in any case, and as the saying goes, deciding to do nothing is itself a decision. Just as an anxious job-interview candidate might well decide to chill out with some deep breaths, full-body stretches, and a power pose, those of us within the scientific community have to make use of whatever ideas are nearby, in order to make the micro-decisions that, in the aggregate, drive much of the directions of science. And, when considering larger ideas, proposals for educational requirements or recommendations for new default statistical or research methods or reorganizations of the publishing system, we need to recognize that our decisions will necessarily rely much more on logic and theory than on direct empirical evidence. This suggests in turn that our reasoning be transparent and openly connected to the goals and theories that motivate and guide our attempts toward fixing science.

References

Gelman, A., and Loken, E. (2012). Statisticians: When we teach, we don't practice what we preach. *Chance* **25** (1), 47–48.

Fiske, S. T., Schacter, D. L., and Taylor, S. E. (2016). Introduction. *Annual Review of Psychology* **67**.